

## ПРИМЕНЕНИЕ МЕТОДА «АНАЛИЗ НАСТУПЛЕНИЯ СОБЫТИЯ (EVENT HISTORY ANALYSIS)» С ПОМОЩЬЮ ПАКЕТА SPSS

А.Я. Бурдяк, магистр экономики, НИСП

Наблюдение за событиями и процессами, развивающимися во времени, неизбежно приводит к сопоставлению этих событий друг с другом, поиску закономерностей и попытке предугадать ход развития процессов. Первые примеры изучения смертности населения в зависимости от возраста датируются 1700-ми гг. [Hald A., 1990]. По мере развития науки разрабатывались разнообразные методы анализа наступления или ненаступления определенных событий, а появление мощной вычислительной техники расширило возможности исследователей в технической обработке информации. На современном уровне исследование процессов и событий, зависящих от времени, проводится в течение, по крайней мере, четырех десятилетий [Кокс Д., Льюис П., 1969. Кокс Д., Оукс Д., 1988] в медицине, демографии, технике, — и область применения подобных исследований расширяется. Модели для изучения длительности какого-либо процесса или времени нахождения в каком-либо состоянии встречаются в литературе под названием моделей «времени жизни» [Магнус Я.Р., Катышев П.К., Пересецкий А.А., 2004] благодаря их изначальному применению в медицине для исследования эффективности лекарственных препаратов при лечении смертельных недугов.

Эконометрические основы методов в данной статье представлены не будут, так как целью изложения является именно прикладной аспект. С теорией можно ознакомиться в учебниках эконометрики, например, по книге Greene [Greene W.H., 1997], содержащей практически все разделы современной эконометрики. Описание существующих типов моделей данного вида и некоторых примеров их применения сделано О. Савинцевой. Лекция по Event History Analysis была прочитана в Независимом институте социальной политики Л.И. Ниворожкиной. Широкому рассмотрению регрессионных методов исследования событий во времени посвящена книга D. Hosmer, S. Lemeshov [David W. et al., 1999], которая удачно объединяет описания нюансов выбора данных, определения и иллюстрации используемых статистических характеристик и интерпретации итоговых результатов моделирования. Курсы по Event History Analysis (другие названия *survival analysis*, *hazard analysis*, *duration analysis*, *failure-time analysis*) включены в программу обучения многих зарубежных университетов, конспекты лекций, задания по курсам и базы данных для выполнения практических заданий доступны в сети Интернет и являются хорошим методическим материалом для исследователей [Scheike Th.]. Недавно опубликована книга данных авторов по этой тематике [Torben Martinussen, Thomas Scheike, 2006]. Особого упоминания заслуживают лекции для социологов, которые читает профессор Кноук в университете Миннесоты [Knoke D.H.], — в его курсе изложение теоретических основ метода проводится в ракурсе практического примене-

ния к задаче исследования возраста первого вступления юношей и девушек в брачный союз. Исследование проводится на базе данных Национального ежегодного обследования молодежи 1979—1991 (National Longitudinal Survey of Youth), осуществляемого Государственным университетом Огайо при финансовой поддержке Министерства труда США.

Так как исследования длительности процессов и наступления событий широко применяются в различных областях, непосредственно не связанных с медициной или демографией, в дальнейшем для метода Event History Analysis будем использовать название Анализ наступления события (АНС), наиболее общее и не связанное с контекстом изучаемых явлений.

АНС проводится в демографических, экономических и социологических исследованиях на микро- и на макроуровнях. Наиболее распространенными примерами применения метода на микроуровне являются:

- возраст завершения образования;
- возраст вступления в самостоятельную жизнь;
- возраст первого вступления в брак;
- длительность супружества;
- возраст женщины на момент рождения ребенка (детей);
- длительность поиска работы;
- динамика повышений по службе;
- частота наступления страховых случаев;
- продолжительность срока службы оборудования в зависимости от условий эксплуатации.

Хотя исторически метод был сконструирован для исследования событий в жизни людей, он может применяться и для анализа хода слияний или закрытий предприятий, перехода на новые технологии, перерастания конфликтов в вооруженные, хода реформ власти или национальных революций.

С помощью метода АНС изучаются данные, протяженные по времени. Они могут быть собраны либо с помощью регулярно повторяющихся опросов одной и той же группы респондентов, либо в результате ретроспективных опросов. Анализируются цензурированные выборки — в модели учитываются и респонденты, для которых интересующее нас событие произошло, и те, для которых оно еще не случилось на момент последнего наблюдения или проведения опроса. В данной статье будет рассмотрен пример применения метода АНС, построенный на данных ретроспективного опроса EES<sup>1</sup>.

## ОПРЕДЕЛЕНИЕ МЕТОДА

Дадим формальное определение: Анализ наступления события (АНС) (Event History Analysis) — *изучение зависимости риска наступления (или ненаступления) события от продолжительности пребывания объекта в группе риска, а также от ряда других характеристик объекта и внешних факторов, влияющих на этот риск.* Хотя из названия метода ясно, что изучается наступление некоторого события, опишем подробнее контекст, в котором применяется метод АНС. Ряд объектов (или респондентов) находится некоторое достаточно длительное время под наблюдением исследователя. Все они без исключения из так на-

<sup>1</sup> Обследование «Образование и занятость» (EES) было проведено Независимым институтом социальной политики (Москва) при финансовой поддержке Научного общества Макса Планка (Германия) весной 2005 г. Концепция и инструментарий обследования были адаптированы к российским условиям Независимым институтом социальной политики с участием Независимой исследовательской группы «Демоскоп» и Института демографических исследований им. Макса Планка (Росток, Германия).

зываемой группы риска, а это означает, что в любой момент с каждым из Объектов может произойти некоторое Событие, при наступлении которого Объект навсегда выбывает из группы риска. (Данное предположение отражает специфику раннего применения метода, разработанного для исследования эффективности лекарственных препаратов при лечении смертельных недугов.) Группа риска является целевой группой нашего исследования. Хорошей иллюстрацией применения метода в технике служит анализ срока службы лампочек накаливания: Объект (лампочка накаливания) находится в группе риска перегореть, и Событие (перегорание лампочки) наступает неизбежно, однократно и безвозвратно.

В пакете SPSS реализованы четыре инструмента для проведения анализа наступления событий: табличный метод (Life Tables), метод Каплана–Мейера (Kaplan–Meier Estimator), регрессия Кокса (Cox Regression) и регрессия Кокса с зависящими от времени объясняющими переменными (Cox Regression with Time Dependent Covariates). В данной работе рассматривается применение метода регрессии Кокса, которую еще называют моделью пропорциональных рисков (*proportional hazard regression model* Cox (1972), Марковские процессы (1960-е гг.)).

Несколько слов о теоретической основе метода. Регрессия Кокса, или модель пропорциональных рисков, — построение риска наступления события как функции, зависящей от времени, и оценка влияния каждой из независимых переменных на этот риск. Риск наступления события — функция от времени — измеряет правдоподобие наступления события в самом ближайшем будущем для тех, кто еще находится в группе риска. Риск наступления события равен предельному значению условной вероятности наступления события во временном промежутке  $[t, t + dt]$  для объектов, еще оставшихся в группе риска на момент времени  $t$ , деленному на длину временного интервала  $dt$ . Риск — не вероятность, и он может быть больше единицы.

Базовые предположения метода АНС следующие:

**(П1) линейность.** Все объясняющие переменные влияют линейно на логарифм функции риска наступления События;

**(П2) независимость объясняющих переменных.** Все объясняющие переменные независимы. В случае присутствия взаимного влияния некоторых регрессоров, в модель должны быть дополнительно включены функции их взаимодействия;

**(П3) риски пропорциональны.** Риски наступления События для любых двух респондентов (или других объектов из группы риска) пропорциональны, и коэффициент пропорциональности не зависит от времени.

На основании предположений **(П1)–(П3)** выбрана функциональная форма модели: регрессия Кокса предполагает, что риск наступления события для  $i$ -того индивида имеет вид:

$$h_i(t) = h_0(t) \times \exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}) \quad (1)$$

или

$$\ln h_i(t) = \ln h_0(t) + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}, \quad (2)$$

где

$h_0(t)$  — «базовый» риск, общий для всех индивидов;

$X_1, \dots, X_p$  — независимые переменные, регрессоры;

$\beta_1, \dots, \beta_p$  — соответствующие коэффициенты.

Базовый риск  $h_j(t)$  — риск наступления события для объекта из референтной группы (для которого все независимые переменные  $X_1, \dots, X_p$  равны нулю). Коэффициенты  $\beta_1, \dots, \beta_p$  отражают влияние каждой из независимых переменных (регрессоров) на функцию риска: при увеличении  $X_j$  на единицу и фиксированных значениях остальных регрессоров, риск наступления события возрастает в  $\exp(\beta_j)$  раз. Необходимо подчеркнуть, что функциональная форма регрессии Кокса (1) отражает базовое предположение модели о пропорциональности рисков (ПЗ): отношение рисков наступления события для любых двух индивидов  $i$  и  $j$  остается неизменным со временем:

$$\frac{h_i(t)}{h_j(t)} = \frac{\exp(\beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip})}{\exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \dots + \beta_p X_{jp})}, \quad (3)$$

то есть риски пропорциональны.

Инструмент регрессия Кокса не предполагает знания или оценок функций распределения переменных-участников модели.

По форме модель пропорциональных рисков похожа на логистическую регрессию, их реализации в пакете SPSS почти идентичны, обе они оцениваются методом максимального правдоподобия.

Метод АНС держится на трех китах: Объекты, События и Время. На языке социологических исследований это Люди (со всеми их характеристиками), События (вернее, одно на всех событие, вероятность наступления которого в жизни каждого человека мы изучаем) и Время (срок от момента попадания в группу риска до момента наступления события). Время — основа метода, оно должно быть определено точно и четко, может измеряться одной из стандартных хронометрических единиц (секундами, днями, месяцами, годами), единственной выбранной для проведения конкретного анализа.

Для применения метода АНС исследователь обязательно должен:

- 1) отобрать целевую группу (группу риска) Объектов для наблюдения;
- 2) узнать о каждом Объекте из целевой группы продолжительность нахождения в группе риска и исход наблюдения: либо наступление События, либо ненаступление События вплоть до завершения наблюдения;
- 3) выделить и собрать информацию о ряде других характеристик и свойств, предположительно влияющих на скорость или риск наступления События в биографии Объекта. Они могут быть постоянными или меняющимися во время наблюдения за Объектом.

Главная задача — изучить, как меняется риск наступления События в зависимости от того, как долго Объект уже находится в группе риска, а также определить насколько важными являются другие его характеристики и как каждая из них влияет на риск наступления события. В социологии наиболее часто используемыми характеристиками являются пол, возраст, регион проживания, состав семьи и пр. Широта выбора объясняющих переменных ограничена только возможностями сбора достоверной объективной информации: например, при изучении формирования предпочтений или склонностей человека важными факторами являются аналогичные характеристики его родителей, культурные традиции, другие особенности уклада жизни; для исследования надежности аппаратуры (ожидаемый срок службы) необходимо учесть сроки и условия предыдущей эксплуатации и т.п. Выбор в руках исследователя, он зависит от целей и задач проекта,

от возможностей и ограничений имеющихся данных и, самое важное, от хорошей интерпретируемости предполагаемых результатов.

## ПРИМЕР ПРИМЕНЕНИЯ МЕТОДА АНС

Перейдем к описанию примера применения метода АНС с помощью регрессии Кокса в пакете SPSS. Изучается длительность поиска работы молодыми специалистами на базе данных обследования EES (2005 г.), охватывающего 6455 респондентов<sup>2</sup>. Объектами нашего исследования являются люди в возрасте старше 18 лет, 1948–1988 годов рождения (год рождения — переменная *ybrn*). В данном случае под Событием мы понимаем устройство респондента на работу. Целевую группу составляют только те респонденты, для которых в возрасте 16 лет обучение еще являлось основным занятием, затем они закончили обучение и через какое-то время основным их занятием стала работа. Таким образом, из анализа исключаются 642 респондента массива EES, которые уже не учились в 16-летнем возрасте, в группе риска остается 5813 человек, на наблюдении за которыми и строится анализ. Группа риска состоит из выпускников учебных заведений, которые не намерены больше учиться, а хотят работать. В данном случае Событие — устройство на работу: как только молодой специалист трудоустроился, он выбывает из группы риска. В случае «блуждания» Объекта между образованием и трудом, мы обязательно должны оговорить которое именно из его многочисленных трудоустройств мы будем учитывать: первое, последнее или первое после получения высшего образования. Так как метод основан на предположении о независимости респондентов друг от друга, повторное появление одного и того же объекта в группе риска НЕ ДОПУСКАЕТСЯ. Кроме информации о точном времени наступления События (или ненаступления) в жизни респондента до момента окончания наблюдения, мы обладаем знанием ряда характеристик о каждом Объекте, которые могут влиять на скорость наступления события. Эти характеристики могут быть как постоянными (пол, цвет кожи), так и меняться за время нашего наблюдения за Объектом (регион проживания, уровень образования, количество детей).

Как только респондент завершает образование, у него появляется шанс устроиться на работу, что в терминологии метода означает, что он попадает в группу риска быть подвергнутым Событию, и в этот момент мы включаем за ним наблюдение (включается счетчик времени, переменная *time* = 0 месяцев). В момент, когда респондент приступает к работе, наступает Событие: переменной *status* присваивается значение 1, сигнализирующее о наступлении события, счетчик времени останавливается, *time* равен количеству месяцев, прошедших с момента завершения обучения (когда был *time* = 0). Если на момент опроса выпускник так и не приступил к работе, счетчик времени тоже останавливается, а переменной *status* присваивается значение 0. В этом случае *time* равен количеству месяцев, прошедших с момента завершения обучения (когда был *time* = 0) до момента окончания наблюдения (момента опроса). Месяц в данном случае является наиболее адекватной единицей измерения времени. Конечно, итоговая картина была бы четче, если бы время поиска работы выражалось в неделях или днях, но имеющиеся в нашем распоряжении данные не позволяют достичь такого совершенства.

<sup>2</sup> Обследование EES проводилось на базе выборки обследования «Родители и дети, мужчины и женщины в семье и обществе» (2004 г.), охватило население в возрасте 19–55 лет из 32 регионов России и репрезентирует население указанных возрастов на национальном уровне.

Таким образом, мы сформулировали задачу и готовы применять метод АНС:

- Событие — трудоустройство молодых специалистов;
- Время — количество месяцев от окончания обучения до начала работы;
- Группа риска — те, кто учился после 16 лет (6455 — 642 = 5813 респондентов);
- Объясняющие переменные — возраст на момент окончания учебы, пол, год рождения.

*Тонкости подбора данных.* При построении данного примера намеренно взяты самые широкие из возможных группы респондентов. Так, можно было бы ограничиться рассмотрением, например, только выпускников в возрасте до 30 лет, или, например, не учитывать тех выпускников, кто не работал после выпуска по ряду объективных причин (армия, отпуск по уходу за детьми) либо учитывать, что они ищут работу только с момента исчезновения перечисленных причин. Углубление в детали, разбиение целевой группы на подгруппы, сравнения и выбор лучшего из возможных результатов — все эти цели постепенно достигаются в ходе дальнейшей работы над моделью, когда уже есть что улучшать и совершенствовать.

При построении модели сначала проверяем, насколько хорошо выполняются предположения метода (П1)–(П3), затем применяем его, интерпретируем и обобщаем результаты. Возможно, в ходе анализа задачу придется переформулировать, добавить новые переменные или подправить зависимости в модели.

На этапе предварительной работы с данными объясняющие переменные необходимо проверить на слабую взаимную коррелированность. Для переменных с интервальной или номинальной шкалой используется коэффициент корреляции Пирсона. Если, по крайней мере, одна из двух переменных имеет порядковую шкалу либо не является нормально распределенной, рассчитывается ранговая корреляция Спирмана. Сила связи двух переменных характеризуется абсолютной величиной коэффициента корреляции. Обычно считается, что значение до 0,2 — очень слабая корреляция, до 0,5 — слабая корреляция, до 0,7 — средняя корреляция, до 0,9 — высокая корреляция, выше 0,9 — очень высокая корреляция. Подробное описание

Рисунок 1. Корреляции (Correlations)

			AGEDU Возраст респондента, когда учеба перестала быть основным занятием	SEXN Пол	YBRN Год рождения
Spearman's rho	AGEDU Возраст респондента, когда учеба перестала быть основным занятием	Correlation Coefficient	1,000	,058(**)	-,067(**)
		Sig. (2-tailed)	.	,000	,000
		N	5813	5813	5813
	SEXN Пол	Correlation Coefficient	,058(**)	1,000	-,082(**)
		Sig. (2-tailed)	,000	.	,000
		N	5813	6455	6455
	YBRN Год рождения	Correlation Coefficient	-,067(**)	-,082(**)	1,000
		Sig. (2-tailed)	,000	,000	.
		N	5813	6455	6455

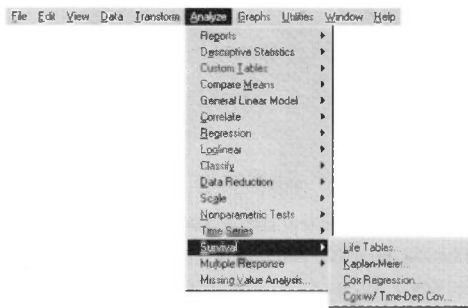
\*\* Correlation is significant at the 0.01 level (2-tailed).

методов обработки данных представлено в пособии по работе в SPSS [Бюль А., Цёфель П., 2005].

В нашем случае три независимых переменных возраст респондента, когда учеба перестала быть его основным занятием *agedu*, пол *sexn* и год рождения *ubrn* очень слабо коррелированы, все корреляции значимы на 1-процентном уровне, и, следовательно, все три указанных переменных одновременно могут быть использованы в качестве независимых переменных в регрессии. Результат корреляционного анализа представлен на *рис. 1*.

Пошагово проиллюстрируем применение метода с помощью SPSS. Для АНС в SPSS встроена процедура регрессии Кокса (Cox Regression) в рубрике Analyze/Survival (*рис. 2*). Запускаем ее.

Рисунок 2

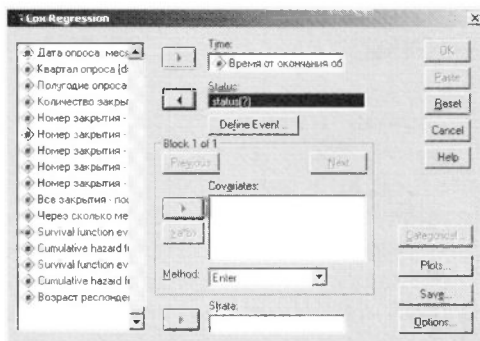


Появляется окно регрессии Кокса (*рис. 3*). Из списка всех переменных, отраженного слева, выбираем заранее подготовленную переменную времени в Time — в нашем случае это *time* «Время от окончания обучения до начала работы», которое измеряется в месяцах. Результатами анализа будут функция риска (шанса) найти работу и ее изменение в зависимости от продолжительности статуса незанятости после окончания образования. Риск меняется в той же временной шкале, что и *time* — ежемесячно.

Переменная в графе Status отражает наступление или ненаступление события на момент завершения наблюдения за данным респондентом. В нашем случае это переменная *status*, и ее построение было подробно описано выше. Следует поместить *status* в соответствующую графу (*рис. 3*).

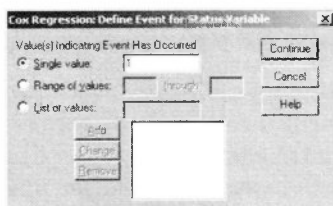
С помощью Define Event мы обязаны указать, какое именно из значений переменной *status* сигнализирует о наступлении события. В нашем

Рисунок 3



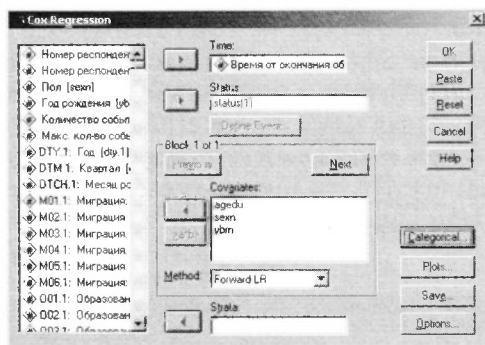
случае это значение  $status = 1$ , указываем это значение (рис. 4). Переменная, сигнализирующая о благоприятном/неблагоприятном исходе наблюдения, не обязана быть создана заранее в виде дамми-переменной со значениями 1 и 0. Сигналом о свершении события может служить интервал значений непрерывной переменной. Например, размер накоплений выше 100 тыс. рублей при страховании вкладов, или пробег автомобиля, например, 40 тыс. км до наступления первого страхового случая. Если мы хотим задать пороговые значения количественной переменной или список значений категориальной переменной статуса, сигнализирующих о наступлении события, вносим их в соответствующие ячейки.

Рисунок 4



Далее, из списка переменных переносим в Covariates все независимые переменные, влияние которых на риск (шанс) наступления события изучаем (рис. 5). Для расчета функции шанса трудоустройства мы используем возраст респондента на момент, когда образование перестало быть его основным занятием *agedu* (участники опроса, которые не учились после 16 лет или еще не завершили образования на время проведения опроса, не входят в группу риска и из анализа исключены); пол *sex* и год рождения *ybrn*. Год рождения включен в модель с целью обнаружения и оценки влияния межпоколенных различий.

Рисунок 5

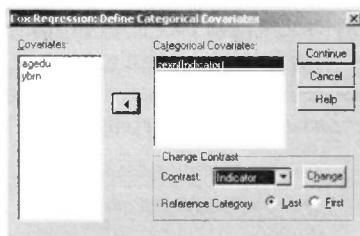


Для учета взаимосвязанных между собой факторов (высоко коррелированных друг с другом) доступен учет в качестве регрессора произведения двух переменных: выделяем из списка слева две переменных и нажимаем  $>a*b>$ . По умолчанию все независимые переменные считаются непрерывными. Если среди них есть категориальные переменные, то для каждой из них необходимо указать *Categorical...* и определить, которая из групп считается референтной (рис. 6). Так, указание для переменной *sex*, что референтная группа Last (в нашем случае это 2 «женский»), пересчи-



тывает переменную вида 1 «мужской»/2 «женский» в дамми-переменную со значениями 1 «мужской»/0 «женский». Та же процедура применяется к категориальным переменным имеющим множество значений. Например *level*, принимающая 7 натуральных значений от 1 до 7 с референтной группой *Last*, будет автоматически переделана в 6 дамми-переменных вида 1/0, эти новые шесть переменных будут включены в регрессию, и коэффициент при каждой из них будет указывать отличие группы респондентов с данным значением *level* от группы респондентов со значением *level* = 7.

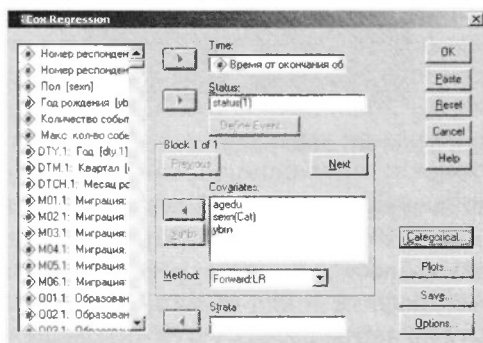
Рисунок 6



Вклады разных значений категориальной переменной не обязательно отличаются на единицу. Условное «расстояние» между категориями может быть представлено в модели другими функциями, предлагаемыми в Contrast.

**Выбор метода.** Предлагается три группы методов (*Method*) оценивания регрессии (рис. 7): метод ввода (*Enter*), пошаговый вперед (*Forward*) и пошаговый назад (*Backward*). Метод ввода оценивает регрессию сразу со всеми независимыми переменными из списка, выдает оценки всех коэффициентов и функцию правдоподобия. При пошаговом «вперед» в регрессию постепенно по одной на каждом шаге дополнительно вводятся независимые переменные в порядке убывания их объясняющей силы, регрессия каждый раз оценивается, и в случае неулучшения объясняющего качества регрессии процесс добора независимых переменных прекращается. При пошаговом методе «назад» мы стартуем с максимального набора объясняющих переменных, постепенно отбрасывая все лишнее и не ухудшая итоговую модель — процедура останавливается на шаге, когда уже остались только хорошо объясняющие функцию риска независимые переменные. Для оценки качества регрессии предлагаются различные статистические процедуры оценивания: тест максимально-

Рисунок 7



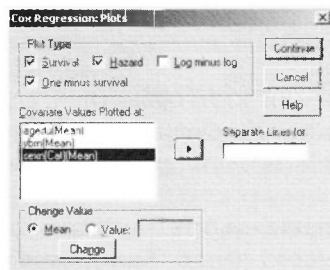
го правдоподобия, тест Вальда и тест Условных вероятностей. В нашем примере была использована пошаговая процедура *ForwardLR* постепенного добавления регрессоров с оценкой качества регрессии с помощью метода максимального правдоподобия (Likelihood Ratio), стандартная и наиболее используемая на первом шаге анализа.

На этом формулировка задачи завершена и по команде Paste в синтаксисе она фиксируется в следующем виде:

```
COXREG
time/STATUS=status(1)
/CONTRAST (sexn)=Indicator
/METHOD=FSTEP(LR) agedu sexn ybrn
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

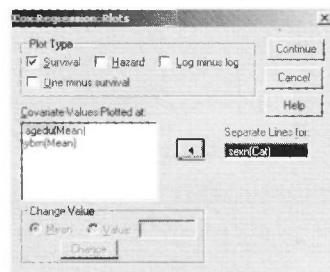
Одновременно с численным оценивая влияния независимых переменных на риск наступления события, можно определить результат графически — в процедуру встроены четыре вида графиков Plots (рис. 8):

Рисунок 8



По умолчанию график рисуется в точке, где каждая из независимых переменных равна своему среднему значению. Если нас интересует график для какого-то конкретного значения переменной, указываем его в *Change Value/Value*. Кроме того, вместо одной линии функции риска (или функции вероятности найти работу) мы можем нарисовать набор кривых для разных категорий. Например (рис. 9), выберем *Separate Lines for* для переменной пол.

Рисунок 9



Тогда в синтаксисе появится строка  
/PATTERN BY sexn

И вместо графика вероятности не найти работу (по-другому она называется функцией «выживания» и равна вероятности того, что Событие не происходит) для всех респондентов, представленной на рис. 10, появятся две линии на графике — для мужчин и для женщин (рис. 11). Видим, что вероятность не найти работу у мужчин выше, чем у женщин.

Рисунок 10

Survival Function at mean of covariates

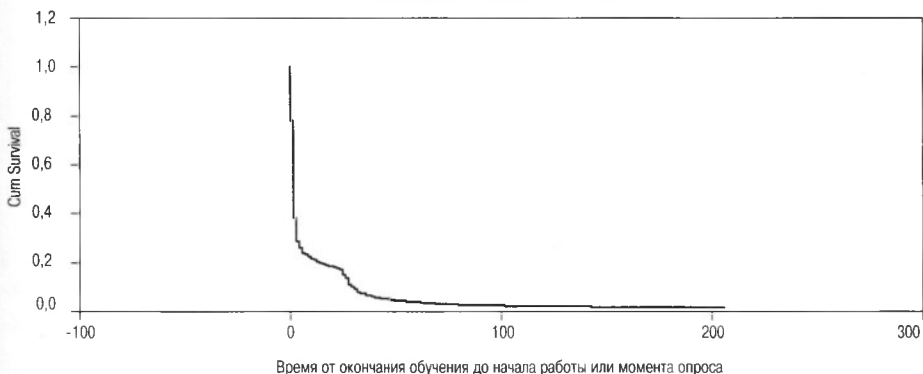
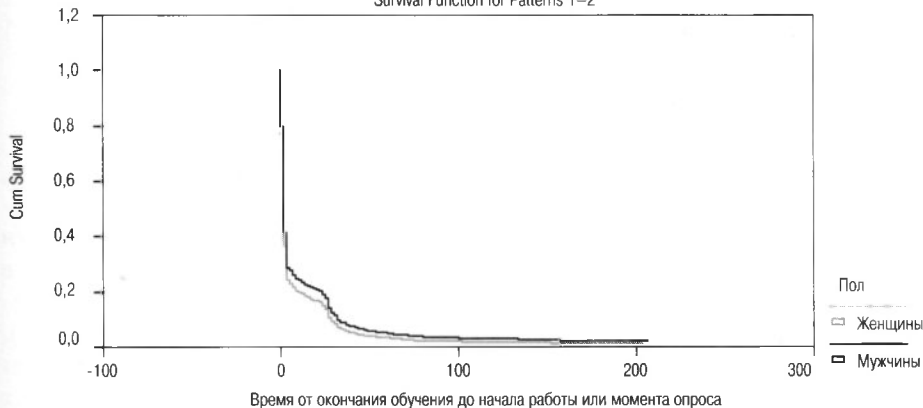


Рисунок 11

Survival Function for Patterns 1-2



Также из графика видно, что функции вероятностей пропорциональны — естественное следствие предположения модели АНС (ПЗ). Численное значение коэффициента пропорциональности, а также остальных параметров регрессии SPSS выдает отдельной таблицей Variables in the Equation (рис. 12). А универсальный критерий коэффициентов модели (Omnibus Tests of Model Coefficients) дает значения функции правдоподобия и критерия 2 для каждого из шагов регрессии. В нашем случае все коэффициенты модели значимы и их значения позволяют численно проинтерпретировать полученный результат. Возраст респондента, в котором он завершает образование и приступает к поиску работы — положительный фактор, каждый дополнительный год возраста молодого специалиста увеличивает его шанс трудоустроиться в 1,036 раз. Так как референтной группой переменной *sex* являются женщины (рис. 13), то отрицательность соответствующего коэффициента  $-0,131$  указывает на то, что риск трудоустроиться после окончания образования для мужчин ниже, чем для женщин.  $Exp(B)$  дает оценку коэффициента пропорциональности: риск мужчин в 0,877 раз ниже, чем риск женщин. Данный результат отражает распространенность явления отложенного трудоустройства женщин после получения образования, например, по причине рождения детей. А мужчина если не начал сразу работать, то и не трудоустраивается в дальнейшем. Более глубокие выводы требуют дальнейшего тщательного анализа имеющихся данных,

Рисунок 12

		B	SE	Wald	df	Sig.	Exp(B)
Step 1	YBRN	-,031	,001	552,052	1	,000	,970
Step 2	AGEDU	,037	,004	71,478	1	,000	1,038
	YBRN	-,029	,001	496,526	1	,000	,971
Step 3	AGEDU	,036	,004	65,820	1	,000	1,036
	SEXN	-,131	,029	21,181	1	,000	,877
	YBRN	-,029	,001	497,153	1	,000	,971

например, исследования других видов занятий, рождения детей и пр., а также рассмотрения полученных результатов в контексте работ других авторов по этой тематике — все это выходит за рамки данной методологической статьи.

Рисунок 13. Кодировка категориальных переменных (Categorical Variable Codings(b))

		Frequency	(1)
SEXN(a)	1=male	2210	1
	2=female	3603	0

a Indicator Parameter Coding

b Category variable: SEXN (Пол)

Переменная *ybrn* дает возможность сравнить между собой поколения. Отрицательность коэффициента указывает на то, что со временем трудоустроиться молодым специалистам становится все труднее. Так, за один год шансы трудоустроиться уменьшаются в 0,971 раз, а за десять лет — в 0,75 раз. То же самое можно выразить в терминах обратных величин: десять лет назад шанс трудоустроиться был в 1,34 раза выше (на 34% выше), чем сегодня. То, что трудоустроиться молодому поколению сложнее, чем это было раньше, — интуитивно ожидаемый результат, так как в условиях плановой экономики существовала строгая система распределения молодых специалистов и проводился контроль за их своевременным трудоустройством.

Оценки соответствующих коэффициентов, а также графики, приведенные в данном тексте, прорисованы для средних значений независимых переменных (рис. 14).

Рисунок 14. Средние значения независимых переменных (Covariate Means and Pattern Values)

	Mean	Pattern	
		1	2
AGEDU	19,634	19,634	19,634
SEXN	,380	1,000	,000
YBRN	1966,406	1966,406	1966,406

Наряду с графиками функций «выживания» для иллюстрации результатов могут использоваться и графики самой функции риска (рис. 15, 16). Поведение функции риска подтверждает сделанные выше выводы о влиянии пола, возраста и года рождения респондента на время поиска им работы. Выбор между функцией выживания и функцией риска зависит от предпочтений исследователя. Чтобы немедленно предоставить право выбора читателю, приведем графики функций риска.

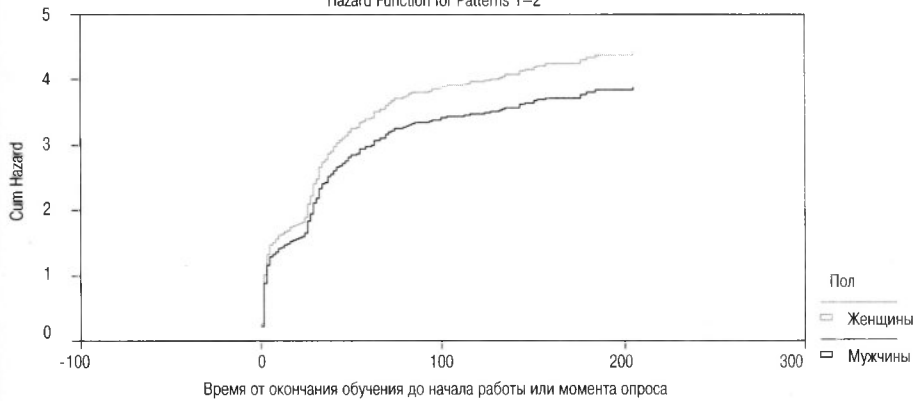
Рисунок 15

Hazard Function at mean of covariates



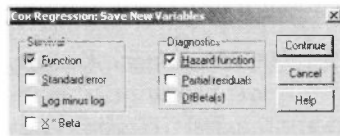
Рисунок 16

Hazard Function for Patterns 1–2



В регрессии Кокса в SPSS можно сохранить значения моделируемых функций (функции риска, вероятность выживания и пр.) для каждого респондента — предусмотрена опция Save, где можно выбрать какие именно значения мы хотим сохранить в массиве данных в качестве предсказанных моделью переменных (рис. 17).

Рисунок 17



Использование данной процедуры позволяет, например, изменить дизайн графиков с помощью других графических редакторов или сделать дополнительные, не предусмотренные самой процедурой регрессии Кокса в SPSS, расчеты.

Метод АНС — простой и удобный способ изучения риска наступления событий, численного оценивания зависимости этого риска от времени и других характеристик объекта и построения прогноза наступления событий. Понимание основных предпосылок и особенностей применения АНС, изложенных в статье, поможет проводить АНС, реализованный в SPSS и

других статистических программах. Данная методика имеет широкую область применения в демографических, экономических и социологических исследованиях, а также с успехом может использоваться на практике в работе страховых компаний и банков.

## ЛИТЕРАТУРА

1. *David W., Hosmer Jr., Lemeshov S.* Applied survival analysis : regression modeling of time to event data. — 1999.
2. *Greene W.H.* Econometric Analysis. 3rd edition. Prentice-Hall, Upper Saddle River. — New Jersey. 1997.
3. *Hald A.* A History of Probability and Statistics and Their Applications Before 1750. — NY: John Wiley & Sons, Inc., 1990.
4. *Knoke D.H.* Course on Event History Analysis, spring 2004, University of Minnesota. <http://www.soc.umn.edu/~knoke/>
5. *Martinussen T.* Introduction to statistics for time to event data, Royal Veterinary and Agricultural University <http://www.matfys.kvl.dk~torbenm/DINA/survival>
6. *Martinussen T., Scheike Th.* Dynamic regression models for survival data // Springer. March 2006.
7. *Scheike Th.* Advanced survival analysis for biostatistical/statistical Ph.D.-students. Faculty of Medicine, University of Copenhagen; course Statistical analysis of survival data 2006 at the Department of Biostatistics. University of Copenhagen. <http://staff.pubhealth.ku.dk/~ts/>
8. *Бююль А., Цёфель П.* SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. — СПб.: ДиаСофтЮП, 2005.
9. *Кокс Д., Льюис П.* Статистический анализ последовательностей событий. — М.: Мир, 1969. *Кокс Д., Оукс Д.* Анализ данных типа времени жизни. — М.: Финансы и статистика, 1988.
10. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс: Учеб. / Изд. 6-е, перераб. и доп. — М.: Дело, 2004.
11. *Савинцева О.* Анализ длительностей до момента прекращения. <http://www.nsu.ru/ef/tsy/ecmr/durat/savinc/savinc.htm>